Title	Note on a generalization of a law of large numbers
Sub Title	
Author	遠藤, 靖(Endo, Yasushi)
Publisher	慶応義塾大学工学部
Publication year	1972
Jtitle	Keio engineering reports Vol.25, No.2 (1972.) ,p.13- 17
JaLC DOI	
Abstract	A sequence of probability distribution functions controling behavior of a sequence of mutually independent random variables is given as an outcome of a "generator" stochastic process. Even if the law of large numbers holds for each outcome of the generator process, the limit may depend on the outcome. In this paper a sufficient condition where the limits in the law of large numbers are constant for almost all outcomes of the generator process is given.
Notes	
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO50001004-00250002-0013

KEIO ENGINEERING REPORTS
VOL. 25 No. 2

NOTE ON A GENERALIZATION OF A LAW OF LARGE NUMBERS

BY
YASUSHI ENDOW

FACULTY OF ENGINEERING
KEIO UNIVERSITY
YOKOHAMA 1972

KEIO ENGINEERING REPORTS
VOL. 25, NO. 2, pp. 13-17, 1972

NOTE ON A GENERALIZATION OF A LAW OF LARGE NUMBERS

YASUSHI ENDOW

Dept. of Administration Engineering, Keio University, Yokohama 223.

(*Received May 15, 1972*)

ABSTRACT

A sequence of probability distribution functions controling behavior of a sequence of mutually independent random variables is given as an outcome of a "generator" stochastic process. Even if the law of large numbers holds for each outcome of the generator process, the limit may depend on the outcome. In this paper a sufficient condition where the limits in the law of large numbers are constant for almost all outcomes of the generator process is given.

§ 1. Introduction

Let us consider Bernoulli trials, i.e., repeated independent trials such that there are only two possible outcomes for each trial, say, " success " and " failure ", and their probabilities remain the same throughout the trials. It is usual to denote the two probabilities by p and $q(=1-p)$. If S_n is the number of successes in n trials, then S_n/n is the frequency of successes and assymptotically tends to p as n increases. Accurately, *as n increases, the probability that the frequency of successes deviates from p by more than any preassigned $\varepsilon>0$ tends to zero, i.e.*

$$P\left\{\left|\frac{S_n}{n}-p\right|<\varepsilon\right\}\longrightarrow i, \text{ as } n\longrightarrow\infty.$$

This is the weak law of large numbers due to J. BERNOULLI (1713). E. BOREL (1909) made a statement much stronger than the Bernoulli's, which has been known as the strong law of large numbers: *For almost every sequence of trials the frequency of successes tends to p, as n increases, i.e.*

$$P\left\{\lim_{n\to\infty}\frac{S_n}{n}=p\right\}=1.$$

A. N. KOLMOGOROV (1928. 1929) obtained a generalization of these two theorems. *If the variances of the random variables* X_k *are bounded by one and the same constant, or more generally if* $\sum_{n=1}^{\infty} \mathrm{Var}(X_n)/n^2 < \infty$, *then the sequence of mutually independent random variables* $X_1, X_2, \cdots$ *obeys the strong law of large numbers, i.e.*

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k \longrightarrow 0, \text{ as } n \longrightarrow \infty$$

holds with probability one.

In this paper we consider one generalization of Kolmogorov's theorem stated above. Suppose that a sequence of probability distribution functions, which governs behavior of a sequence of mutually independent random variables (abbreviatied a seq. of mirv.'s) is not given *a priori* but is determined by an outcome of a stochastic process called, a generator process. Even if the law of large numbers holds for each outcome of the generator process, the limit generally depends on the outcome. In this paper a sufficient condition imposed on the generator process is derived so that the limits are the same for almost all outcomes of the generator process.

§ 2. The main theorem.

Let us introduce some notations and definitions.

Let $I=\{1, 2, 3, \cdots\}$, $I_0=\{0, 1, 2, \cdots\}$, $R=(-\infty, \infty)$ and $S=\{F_i, i\in I\}$, where F_i's are probability distribution functions. Put $W_1=(I)^{I_0}$ and $W_2=(R)^{W_1}$. Simbols BW_1 and BW_2 denote the smallest σ-algebras containing cylinder subsets of W_1 and Borel cylinder subsets of W_2 respectively. Here we define two shift-transformations T_1: $W_1 \to W_1$ and T_2: $W_2 \to W_2$ by

$$W_1 \ni w_1 = (i_0, i_1, i_2, \cdots) \longrightarrow T_1 w_1 = (i_1, i_2, i_3, \cdots) \in W_1,$$

and

$$W_2 \ni w_2 = ((x_1^{(0)}, x_2^{(0)}, \cdots), (x_1^{(1)}, x_2^{(1)}, \cdots), \cdots) \longrightarrow$$
$$T_2 w_2 = ((x_1^{(1)}, x_2^{(1)}, \cdots), (x_1^{(2)}, x_2^{(2)}, \cdots), \cdots) \in W_2$$

respectively.

Now let us introduce a probability measure P_1 on a measurable space (W_1, BW_1) and define $X_n(\cdot)$ on W_1 as a coordinate function, i.e. $X_n(w_1)=i_n$ for $w_1=(i_0, i_1, \cdots, i_n, \cdots)$. Let us call the sequence of random variables $\{X_n(\cdot): n\geqslant 0\}$ a generator process. Note that if T_1 is measure-preserving, then for every $i\in I$,

$$P_1\{w_1: X_n(w_1)=i\} = P_1\{w_1: X_0(w_1)=i\}, \quad \text{for} \quad n\geqslant 0.$$

Let P_2 be a probability measure on (W_2, BW_2) satisfying the condition that for every integers $k, n_1<n_2<\cdots<n_k, i_1<i_2<\cdots<i_k$ and reals $x_1, x_2, \cdots, x_k$,

$$P_2\{w_2: Y_{n_1}(i_1, w_2)\leqslant x_1, Y_{n_2}(i_2, w_2)\leqslant x_2, \cdots, Y_{n_k}(i_k, w_2)\leqslant x_k\}$$
$$= F_{i_1}(x_1)F_{i_2}(x_2)\cdots F_{i_k}(x_k),$$

where $F_{i_j} \in S$ and $Y_{n_j}(i_j, \cdot)$ is a coordinate function on W_2 such that

$$Y_{n_j}(i_j, w_2) = x_{i_j}^{(n_j)}$$

for $w_2 = ((x_1^{(0)}, x_2^{(0)}, \cdots), \cdots, (x_1^{(n_j)}, x_2^{(n_j)}, \cdots), \cdots) \in W_2$. Thus $\{Y_n(i, \cdot): n \geqslant 0, i \geqslant 1\}$ is a seq. of mirv.'s on (W_2, BW_2, P_2). Note that the shift-transformation T_2 is measure-preserving and mixing. Let (W, BW, P) be a product probability space of (W_1, BW_1, P_1) and (W_2, BW_2, P_2). We define a seq. of r.v.'s $\{Z_n(\cdot): n \geqslant 0\}$ on (W, BW, P) and a transformation $T: W \rightarrow W$ such that

$$Z_n(w) = Z_n(w_1, w_2) = Y_n(X_n(w_1), w_2)$$

and

$$Tw = T(w_1, w_2) = (T_1 w_1, T_2 w_2)$$

respectively.

Under the assumptions stated above, we have the following theorem.

THEOREM. *If T_1 is measure-preserving and ergodic, and if $\sum_{i \in I} |m_i| p_i < \infty$, then for a.a. w (almost all w),*

$$\frac{1}{n} \sum_{k=0}^{n-1} Z_k(w) \longrightarrow EZ_0 = \int_W Z_0(w) P(dw), \quad as \quad n \longrightarrow \infty,$$

where

$$m_i = \int_{-\infty}^{\infty} x dF_i(x) \quad and \quad p_i = P_1\{w_1: X_0(w_1) = i\} \quad for \quad i \in I.$$

Proof. From measure-preservation of T_1 and T_2 it follows that for every $n \geqslant 0$, $E|Z_n| = \sum_{i \in I} E|Y_n(i, w_2)| P_1\{w_1: X_n(w_1) = i\} = \sum_{i \in I} |m_i| p_i < \infty$. Thus by Birkhoff's individual ergodic theorem (c.f. [3]) there exists a T-invariant and integrable function $\hat{Z}(w)$ such that, for a.a. w

$$\frac{1}{n} \sum_{k=0}^{n-1} Z_k(w) \longrightarrow \hat{Z}(w), \quad \text{as} \quad n \longrightarrow \infty,$$

and $EZ_k = E\hat{Z}$ for $k \geqslant 0$. Since T_1 is ergodic and T_2 is mixing, T is ergodic. This can be proved in the same way as the mixing theorem (c.f. [3]).

It follows from this that $\hat{Z}(w) = \text{const.}$ for a.a. w. Therefore $EZ_0 = E\hat{Z} = \text{const.} = \hat{Z}(w)$ for a.a. w and this completes the proof.

COROLLARY. *If the same assumptions as in the theorem are satisfied and if also $E|Z_0|^2 < \infty$, then*

$$E\left|\frac{1}{n} \sum_{k=0}^{n-1} Z_k(w) - EZ_0\right|^2 \longrightarrow 0, \quad \text{as} \quad n \longrightarrow \infty.$$

Proof. Since T is measure-preserving, we have $E|Z_n|^2 = E|Z_0|^2 < \infty$ for $n \geqslant 0$. It follows from this by the mean ergodic theorem due to J. von NEUMANN (1932) that there exists a T-invariant random variable $Z^*(w)$ such that

$$E\left|\frac{1}{n} \sum_{k=0}^{n-1} Z_k(w) - Z^*(w)\right|^2 \longrightarrow 0, \quad \text{as} \quad n \longrightarrow \infty.$$

This implies that as $n \to \infty$,

$$\frac{1}{n} \sum_{k=0}^{n-1} Z_k(w) \longrightarrow Z^*(w)$$

in probability. Choosing a subsequence $\{n^*\} \subset \{n\}$, we have that for a.a.w

$$\frac{1}{n^*} \sum_{k=0}^{n^*-1} Z_k(w) \longrightarrow Z^*(w), \quad \text{as} \quad n^* \longrightarrow \infty.$$

Hence by the theorem stated above, we must have $Z^*(w) = \hat{Z}(w)$ a.e. This completes the proof.

§ 3. Example

Consider a Markov chain as a generator process. Let $I = \{0, 1, 2, \cdots, N\}$, $I_0 = \{0, 1, 2, \cdots\}$, $R = \{0, 1\}$ and $S = \left\{F_i = \left(\frac{i}{N}, 1 - \frac{i}{N}\right), i \in I\right\}$. We interpret (W_1, BW_1) and (W_2, BW_2) as the same sense in § 2. Let also

$$M = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \cdot \\ \cdot \\ \cdot \\ N-2 \\ N-1 \\ N \end{array} \overset{\begin{array}{cccccccccc} 0 & 1 & 2 & 3 & \cdot & \cdot & \cdot & N-2 & N-1 & N \end{array}}{\begin{pmatrix} 1-c & c & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 1-c & 0 & c & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1-c & 0 & c & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1-c & 0 & c & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & 0 & 1-c & 0 & c & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 1-c & 0 & c & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1-c & 0 & c & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1-c & 0 & c \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1-c & c \end{pmatrix}}, \tag{14}$$

where $0 < c < 1$.

A normalized non-negative solution of the system of equations

$$vM = v \tag{15}$$

is $p = (p_i \colon 0 \leqslant i \leqslant N)$, where

$$\begin{aligned} p_i &= \frac{1+r}{1-r^{N+1}} r^i, \quad \text{when} \quad r \neq 1 \\ &= \frac{1}{N+1}, \quad \text{when } r = 1, \end{aligned} \tag{16}$$

and

$$r = \frac{c}{1-c}.$$

Let $\{X_n(\cdot): h \geqslant 0\}$ be a Markov chain, which is defined by the initial distribution p and the transition matrix M, on (W, BW, P_1) with the state space S. It follows from $pM=p$ that a shift-transformation T_1: $W_1 \to W_1$ defined by

$$X_n(T_1 w_1) = X_{n+1}(w_1) \qquad (n \geqslant 0)$$

is measure-preserving and ergodic. Let P_2 be a probability measure on (W_2, BW_2) satisfying the condition that for every integers $k, n_1 < n_2 < \cdots < n_k, i_1 < i_2 < \cdots < i_k$ and $e_1, e_2, \ldots, e_k$

$$P_2\{w_2\colon Y_{n_j}(i_j, w_2) = e_j,\ 1 \leqslant j \leqslant k\} = \prod_{j=1}^{k} X(i_j, e_j),$$

where $X(i_j, e_j) = i_j/N$, if $e_j = 1$ and $1 - i_j/N$, if $e_j = 0$, and $Y_{n_j}(i_j, \cdot)$ is a coordinate function on W_2. Transformations T_2 and T, a probability space (W, BW, P) and a seq. of r.v.'s $\{Z_n(\cdot): n \geqslant 0\}$ are understood to be defined in the same way as in § 2. Then by the theorem in § 2, as $n \to \infty$, for a.a.w

$$\frac{1}{n} \sum_{k=0}^{n-1} Z_k(w) \longrightarrow \sum_{i \in I} p_i \frac{i}{N} = \left(\frac{r(1-r^N)}{N(1-r)^2} - \frac{r^{N+1}}{1-r} \right) \frac{1+r}{1-r^{N-1}}, \text{ when } r \neq 1$$

$$= \frac{1}{2}, \text{ when } r = 1.$$

This relation also holds in the mean convergence sense as justified by the corollary to the theorem stated above.

Acknowledgment

The auther wishes to express his gratitude for the valuable comments received from Prof. Heihachi SAKAMOTO at Keio Univ. He also wishes to express his thanks to Prof. Gisiro MARUYAMA at Tokyo Univ. of Education for the help with improvements of this paper.

REFERENCES

CHUNG, K. L., (1966): *Markov chains with stationary transition probabilities*, Springer, 92–93.
GNEDENKO, B. V., (1969): *The theory of probability*, Mir Pub., 219.
HALMOS, P. R., (1956): *Lectures on ergodic theory*, Math. Soc. Jap., 13–21, 39.
LOÈVE, M., (1963): *Probability theory*, Van Nostrand, 14–26.